Change My View: How Rhetorical Strategy Modulates Rational Belief Change

David Freeborn
Northeastern University London

Malihe Alikhani
Khoury College of Computer Sciences
Northeastern University

Anthony Sicilia
West Virginia University

Abstract

Bayesian convergence theorems suggest that agents who share evidence should eventually agree, yet multi-belief Bayesian models show how shared evidence can rationally polarise agents whose higher-order priors diverge. I argue that effective persuasion under such conditions requires higher-order alignment: convergence on the background assumptions through which evidence is weighted. A formal model links mediating-node alignment to reduced posterior disagreement, and evidence from 3,051 r/ChangeMyView debates supports the predicted pattern: concession and empathy positively predict belief revision, while direct challenge and deflection negatively predict it. The result reframes concession as an epistemic strategy.

1 Introduction

On 27 September 2018, the United States watched Christine Blasey Ford and Brett Kavanaugh testify before the Senate Judiciary Committee. Pre-hearing polling found

Preprint prepared for PSA 2026 Please cite the published version if available.

the public largely undecided: 42% were unsure whom to believe, while belief in Ford (32%) and Kavanaugh (26%) was roughly balanced (Marist Poll, 2018b). Days later, after exposure to the same sworn testimony, the undecided share collapsed to 22% and opinion polarised sharply along partisan lines: 73% of Democrats believed Ford; 74% of Republicans believed Kavanaugh (Marist Poll, 2018a). The very event designed to produce shared judgment instead drove the poles further apart.

This episode illustrates a tension that runs through contemporary epistemology. Bayesian convergence theorems guarantee that agents with absolutely continuous priors who share evidence must eventually agree (Blackwell and Dubins, 1962). Aumann's agreement theorem shows that common-knowledge posteriors under common priors must be identical (Aumann, 1976). Yet the Kavanaugh case, and countless lower-stakes exchanges, shows shared evidence generating divergent updates. How are we to reconcile these formal guarantees with the frequency of persistent disagreement?

Recent formal work helps us to see how this is possible. When agents hold *multiple, probabilistically entangled beliefs*, Bayesian updating on shared evidence can rationally generate persistent polarisation (Jern et al., 2014; Freeborn, 2023, 2024). Evidence does not speak for itself; its impact on a target proposition is mediated by background credences (about source reliability, ideological priors, the weight of testimony) that differ across agents. Two fully rational Bayesians can therefore update on identical data and move further apart.

This helps us to reframe the puzzle of persuasion. If divergence stems from misaligned higher-order priors rather than irrationality, then successful persuasion must involve more than presenting good evidence. It must also *align* the background assumptions through which evidence is processed. The central thesis of this paper is:

> *Under conditions of rational polarisation driven by entangled higher-order priors, effective persuasion depends on first aligning the mediating assumptions through which evidence is assessed. Concession and empathy are candidate mechanisms for achieving such alignment.*

I defend this thesis in three steps. First, I introduce the relevant formal tools (Bayesian multi-belief networks and Stalnaker's common ground) and present a result showing that reducing divergence on mediating nodes reduces posterior disagreement on the target (§2). Second, I present computational evidence from 3,051 debates on Reddit's r/ChangeMyView, finding a robust empirical pattern consistent with the alignment hypothesis and difficult to reconcile with a simple "more evidence is better" picture (§§3–4). These results are independently interesting as data about the dynamics of persuasion; they also resonate with the formal model in ways I develop in

the philosophical discussion (§5), where I draw implications for the conciliationism debate and for the relationship between dialectical and rhetorical persuasion.

2 Background and Formal Tools

2.1 Two Traditions: Dialectical and Rhetorical Persuasion

Philosophers and theorists have long distinguished between dialectical and rhetorical approaches to persuasion. The dialectical tradition, rooted in Aristotle's *Topics* and developed in contemporary form by Walton (1995) and pragma-dialecticians such as van Eemeren and Grootendorst (2004), views persuasion as fundamentally truth-oriented: through reasoned argument and shared evidence, interlocutors should converge on the correct view. This perspective finds formal expression in Bayesian convergence results. The rhetorical tradition, by contrast, emphasises how speakers actually influence audiences. Classical rhetoric (Aristotle's *Rhetoric*) identifies three modes of persuasion: *logos* (logical argument), *ethos* (credibility of the speaker), and *pathos* (emotional appeal). Contemporary work in this tradition, including Mercier and Sperber (2011)'s influential argumentative theory of reasoning, argues that reasoning evolved not primarily to track truth in isolation but to persuade others within cooperative social contexts.

These two traditions sometimes pull in different directions: what *ought* to persuade a fully rational agent need not always work in practice, and what *does* work need not always involve the best evidence. The analysis I offer here suggests a partial reconciliation. Certain rhetorical strategies, specifically concession and empathy, may serve a distinctly *epistemic* function by preparing the inferential conditions under which dialectical reasoning becomes effective. I do not claim to have fully established this reconciliation, but the formal model and empirical evidence below offer grounds for taking it seriously.

2.2 Common Ground

Before turning to the formal model, it helps to introduce the notion of common ground. Stalnaker (2002) models conversation through a *context set*: the set of possible worlds compatible with what both parties accept. Assertion aims to eliminate worlds from this set; successful communication shrinks the context set toward the truth. Clark and Brennan (1991) elaborates: conversation is a cooperative game whose players continually update their common ground.

For present purposes, the key point is that common ground represents the set

of propositions over which interlocutors' beliefs are sufficiently aligned. Expanding common ground (by accepting propositions the other party holds) creates shared footing on which further argument can proceed. Conversely, when speakers *contract* common ground (by contesting a proposition the other party took for granted), they may force the interlocutor into a defensive posture that makes subsequent evidence harder to accept. Sperber et al. (2010) describe this in terms of *epistemic vigilance*: audiences maintain standing filters calibrated to the perceived reliability and intentions of the speaker. A speaker who signals shared assumptions lowers those filters; one who opens with confrontation raises them. This concept will prove central to interpreting both the formal model and the empirical results.

2.3 Multi-Belief Bayesian Networks and Rational Polarisation

Following Jern et al. (2014) and Freeborn (2023, 2024), represent each agent's belief state as a Bayesian network: a directed acyclic graph whose nodes are propositions and whose edges encode conditional dependencies. Higher nodes represent background beliefs (ideological commitments, reliability assessments, evaluative standards); lower nodes represent first-order claims.

Two agents may share the Bayesian network's *structure* while differing in *prior assignments* over higher-order nodes. When they do, shared evidence can rationally push their posteriors on a target proposition in opposite directions: evidence is informative about the target only via mediating nodes, and if agents weight those mediating nodes differently, the same datum receives divergent interpretations. Jern, Chang, and Kemp demonstrate that such Bayesian belief polarisation is "not only possible but relatively common"; Freeborn extends this to show that persistent polarisation and even predictable factionalisation can arise under full Bayesian rationality.

This result bears on the debate between conciliationist and steadfast responses to peer disagreement. Conciliationists hold that discovering a peer's disagreement should lead one to revise one's confidence substantially (Christensen, 2007; Elga, 2007). Steadfast views maintain that one's total first-order evidence can justify standing firm (Kelly, 2008). The multi-belief model suggests a more nuanced picture: whether conciliation is appropriate depends on *where* in the Bayesian network the disagreement originates. Disagreement rooted in different first-order evidence may warrant steadfastness; disagreement rooted in divergent higher-order priors calls for a different kind of intervention, one that targets the mediating nodes themselves.

2.4 A Formal Result: Alignment and Convergence

Consider a minimal Bayesian network with three nodes: $E \to M \to P$. Let P be a target proposition, M a binary mediating node (e.g., "the source is reliable"), and E a piece of shared evidence. Two agents, A and B, share this structure and agree on the conditional distributions $\Pr(P \mid M)$ and $\Pr(M \mid E)$, but differ on the prior $\Pr(M)$: agent A assigns $\Pr_A(M) = h$ and agent B assigns $\Pr_B(M) = l$, where $h > l$.

After observing E, each agent's posterior on P is:

$$\Pr_i(P \mid E) = \Pr(P \mid M)\ \Pr_i(M \mid E) + \Pr(P \mid \neg M)\ \Pr_i(\neg M \mid E) \tag{1}$$

Define $\delta = [\Pr(P \mid M) - \Pr(P \mid \neg M)]$, the evidential leverage of M on P. The posterior gap on P equals $\delta \cdot |\Pr_A(M \mid E) - \Pr_B(M \mid E)|$. Since the posterior gap on M inherits the ordering of the prior gap (both agents update on the same likelihood ratio), reducing $|h - l|$ *guarantees* a proportional reduction in posterior disagreement on P. This holds regardless of the direction or strength of E; the result is purely structural.

The simplicity of this model is both its strength and its limitation. Real belief networks have many mediating nodes, and a concession that aligns one may activate disagreement on another. Multi-step conversations involve sequential updates where alignment and evidence alternate. The minimal model does not capture these dynamics. What it does capture is the core mechanism: that the gap on P is structurally downstream of the gap on M, so interventions on M are not mere rhetoric but have formal consequences for how all subsequent evidence is processed. Richer models would be needed to study the sequencing and interaction effects that real conversations involve, but the basic directional prediction (alignment facilitates convergence, activation of contested nodes hinders it) should generalise.

Linking this to common ground: a concession (publicly accepting a proposition the interlocutor holds) adds that proposition to the common ground in Stalnaker's sense. In the Bayesian network, this corresponds to narrowing the prior gap on a mediating node, with the consequence just derived: subsequent evidence will produce more convergent updates on the target. Empathy performs a parallel function at the level of evaluative attitudes, signalling shared higher-order commitments (values, standards, trust) and thereby reducing what Sperber et al. (2010) call *epistemic vigilance*: standing filters against unreliable or adversarial communicators.

This generates a testable prediction: conversational moves that expand common ground should facilitate subsequent belief change, while moves that activate contested mediating nodes should hinder it.

3 Computational Evidence

Reddit's r/ChangeMyView (CMV) provides a useful setting for studying persuasion dynamics. The setting differs from formal philosophical thought experiments in being messy, informal, and large-scale, but it shares the core structure that the model describes: agents with entangled priors exchange arguments, and belief change is publicly signalled. I use CMV as a *general test* of the alignment hypothesis's predictions.

3.1 Data and Design

The dataset is the "Winning Arguments" corpus of Tan et al. (2016): 3,051 original posts, 293,297 comments, 34,911 speakers. When a reply changes an original poster's (OP's) mind, the OP awards a *delta* (Δ). This binary signal almost certainly understates subtler shifts, so I treat all findings as conservative. CMV users self-select for good-faith engagement and English fluency; politically polarising topics are unevenly represented.

Following Sicilia et al. (2024), each thread is truncated at a random intermediate turn K, and a large language model (Llama-3, 8B and 70B, instruction-tuned) forecasts whether a delta will eventually be awarded. The model's 1–10 rating is mapped to a probability $p = (k-1)/9$. This baseline captures *content-level* cues but has no explicit representation of rhetorical strategy, providing a principled comparison: do explicitly coded strategies add predictive power beyond what conversational content already provides?

3.2 Strategy Annotation

A separate LLM pipeline identifies rhetorical strategies in each reply. First, free elicitation extracts verb-phrase strategy descriptions ($\sim$1,700 unique phrases). Second, clustering and author review produce a ten-category taxonomy: *conceding/compromising*, *building empathy/rapport*, *establishing credibility*, *logical reasoning/evidence*, *challenging assumptions*, *alternative perspectives*, *emotional appeals*, *reframing/redefining*, *setting boundaries*, and *deflecting/diverting*. A parallel taxonomy derived from the social-influence literature yields convergent results; I report the LLM-induced taxonomy, which shows tighter clustering.

Each reply is coded as a 10-element binary vector $\mathbf{s}$. Two features of this design deserve comment. First, the strategy labels are LLM-generated with limited human validation, introducing noise and possible systematic bias. Manual inspection of a

random subsample of 200 replies suggested high face validity for the concession and challenge labels, and moderate validity for empathy, but a formal inter-annotator study is ongoing. In particular, the category "logical reasoning and evidence" may capture *mentions* of evidence rather than dialectically effective deployment, a distinction the LLM pipeline is not designed to draw. I take the labels as noisy but informative proxies, and note that noise should attenuate rather than inflate the observed associations. Second, using LLMs for both forecasting and annotation raises a circularity concern. The mitigation is structural: the baseline forecaster receives only raw content, while labels are extracted by a separate pipeline with different prompts. The regression tests whether *decomposing* the signal into named categories adds power beyond the raw content.

3.3 Regression Models

Three logistic models relate strategy use to delta outcomes:

- **Strategy-only (S):** $\text{logit}(y) = \boldsymbol{\beta}^\top \mathbf{s}$. Tests whether strategy labels alone predict persuasion.
- **Additive (LLM + S):** adds the baseline forecast p as a covariate.
- **Interaction (LLM $\times$ S):** includes the element-wise product $p\mathbf{s}$, allowing strategy efficacy to vary with conversational context.

All coefficients are estimated via maximum-likelihood with ten-fold cross-validation at the conversation level.

4 Results

Llama-3 70B achieves a Brier score of 0.24 (F1 = 0.78) after uncertainty calibration, up from 0.31 (F1 = 0.56) without. The interaction model consistently outperforms both the strategy-only and additive models, suggesting that strategy efficacy is context-dependent. Table 1 reports coefficients for the interaction model.

The two strategies most closely associated with *expanding common ground* (concession and empathy) are the only significant positive predictors. The strategies most likely to *activate contested mediating nodes* (direct challenge and deflection) are the strongest negative predictors. "Establishing credibility" is non-significant (+0.07, $p = 0.298$): credibility appeals target the speaker's status rather than shared evaluative assumptions, and so neither expand common ground nor activate contested nodes.

Table 1: Logistic regression coefficients (interaction model, LLM-induced taxonomy).

Strategy	**Coeff.**	p**-value**
Positive predictors ($p < 0.05$)		
Conceding and compromising	$+0.86$	< 0.001
Building empathy and rapport	$+0.39$	< 0.001
Non-significant		
Establishing credibility & authority	$+0.07$	0.298
Emotional appeals	$+0.02$	0.805
Setting boundaries & limitations	-0.10	0.519
Reframing and redefining	-0.11	0.086
Providing alternative perspectives	-0.14	0.157
Negative predictors ($p < 0.05$)		
Using logical reasoning & evidence	-0.16	0.029
Challenging assumptions & arguments	-0.37	< 0.001
Deflecting and diverting attention	-0.48	< 0.001

The negative coefficient on logical reasoning (-0.16) requires cautious interpretation. Given the annotation limitations noted in §3.2, this coefficient may reflect evidence deployed *without* prior alignment rather than a genuine failure of logic. I do not wish to overstate this finding; what can be said is that the category as labelled is a negative rather than positive predictor, which is at least *surprising* on a simple evidentialist picture.

This pattern is difficult to explain on a "more evidence is better" model, which would predict logical reasoning and evidential challenges as positive predictors. It is consistent with the alignment hypothesis. But I should be clear about what the data do and do not show. They show a robust predictive pattern. They do not by themselves establish that concession works *by* aligning higher-order priors; alternative explanations, including politeness norms, reply length, and topic effects, cannot yet be fully excluded. What makes the pattern interesting is not just the correlations in isolation, but the fact that the formal model gives us reason to *expect* exactly this pattern, and that a simple evidentialist model does not.

5 Philosophical Discussion

5.1 Concession and the Alignment Mechanism

The formal result from §2.4 provides a framework for interpreting the regression coefficients, though the interpretation remains suggestive rather than conclusive. When a speaker concedes a point, they publicly accept a proposition the interlocutor already holds. In Stalnaker's terms, this adds a proposition to the common ground; in the Bayesian network, it narrows the prior gap on a mediating node, with the formal consequence that subsequent evidence produces more convergent updates.

Consider a concrete illustration from the corpus.[1] An OP argues that earning a high salary at an unpleasant job is preferable to earning less at an enjoyable one. Two challengers respond. The unsuccessful reply opens: "The problem with that line of thinking is that you don't properly factor happiness into your cost-benefit analysis. . . " The successful reply, whose thread eventually earns a delta, opens differently: "Beyond a point you are right. If I moved from a fun job at $25K to a dull one at $70K, I'd probably be happier too. But after reaching a level of comfort, extra money matters less and loving your work matters more." The respondent then offers personal experience of turning down a higher-paying offer.

On the alignment reading, the contrast is clear. The unsuccessful reply opens by attacking the OP's reasoning framework ("the problem with that line of thinking"), contesting precisely those mediating nodes on which the OP's priors are strongest: confidence in their own cost-benefit judgment. The successful reply begins by conceding on a mediating node, validating the OP's core intuition (more money *does* improve wellbeing) within a specified range, then introducing a threshold effect that reframes the question. The concession narrows the inferential gap: because the OP and the respondent now agree that money matters up to a point, the subsequent evidence (personal experience of declining a raise) is routed through a shared channel rather than being dismissed as coming from someone who "doesn't understand" the value of money.

Empathy operates at the level of evaluative attitudes rather than factual propositions. By signalling shared concerns (acknowledging the interlocutor's stakes, validating their framing of the problem), the speaker aligns higher-order nodes that represent not *what is true* but *what matters* and *whom to trust*. In terms of epistemic vigilance (Sperber et al., 2010), empathetic signals reduce the prior probability the interlocutor assigns to "this speaker is adversarial or untrustworthy," a mediating

[1] Lightly paraphrased to conserve space; original thread available in the Winning Arguments dataset of Tan et al. (2016).

node that, when activated, discounts all subsequent evidence regardless of its quality.

On this reading, aligning higher-order nodes is not mere social lubrication but is *epistemically consequential*: it changes the inferential pathways through which subsequent evidence is processed.

5.2 Conciliationism Refined

The multi-belief model suggests that the conciliationism debate may benefit from a structural refinement. The standard debate asks a single question: when you discover that an epistemic peer disagrees, should you revise your credence? Conciliationists like Christensen (2007) and Elga (2007) answer yes, substantially; steadfast theorists like Kelly (2008) answer that your first-order evidence can justify holding firm. But both sides frame the question in terms of credence in a single proposition. The multi-belief model reveals that this framing obscures a crucial structural distinction.

The question is not simply whether to concede when faced with peer disagreement, but *where* in the Bayesian network to concede. Wholesale conciliation (splitting the difference on the target proposition itself) may be unnecessary and even counterproductive if the disagreement on P is downstream of an identifiable higher-order divergence. What is needed is targeted concession: alignment on the mediating nodes that govern how evidence flows, followed by the introduction of first-order evidence through the now-shared channels.

This is neither the conciliationist's blanket softening nor the steadfast theorist's stubborn confidence, but a third option: strategic, structurally informed concession directed at the epistemic bottleneck. The formal result makes this precise: the posterior gap on P is proportional to the gap on M, scaled by the evidential leverage δ. Conciliation *on* M has maximum impact; conciliation directly on P ignores the structural source of disagreement. A speaker who concedes "you're right that money improves wellbeing" and then introduces a threshold is conciliating on exactly the right node; a speaker who simply splits the difference on the target ("maybe high-paying unpleasant jobs are only *somewhat* preferable") conciliates on the wrong one.

The CMV data offer indirect support. The interaction model's superior fit indicates that concession is most predictive of belief change in conversations where the baseline probability of persuasion is moderate (i.e., where mediating assumptions are contested but not hopelessly divergent). In hopeless cases (very low baseline p), even concession cannot bridge the gap; in easy cases (high baseline p), it is unnecessary. This conditional pattern is what the alignment hypothesis predicts.

5.3 Rival Explanations

Before turning to further implications, it is worth considering the strongest alternative explanations for the observed pattern. First, one might propose that concession and empathy simply track *politeness*, and that polite people are more persuasive for social rather than epistemic reasons. This cannot be fully excluded, but it does not explain the negative coefficient on challenge: direct challenge is often perceived as rude, yet simple politeness would predict that *credibility appeals* (a socially positive strategy) should be a strong positive predictor; they are not. Second, one might argue that the results reflect *reply length or effort*: concession and empathy may correlate with longer, more thoughtful replies. This is plausible as a partial explanation, but it does not explain why the interaction model outperforms the additive one, since reply length should not interact with the baseline forecast in the way that strategy does. Third, one might worry about *topic effects*: perhaps concession is more common in threads about topics where minds are easily changed. Topic controls are needed to rule this out, and I flag this as an important direction for future work.

None of these alternatives is decisively refuted by the present data. But the alignment hypothesis does something the alternatives do not: it unifies the positive and negative coefficients under a single mechanism and connects them to independently motivated formal results about Bayesian belief revision.

5.4 Returning to Kavanaugh

The Ford-Kavanaugh hearings now appear as a large-scale instance of the entrenchment mechanism. The testimony was a maximally confrontational evidential event: two incompatible accounts, each supported by emotional intensity, each activating precisely those mediating nodes (beliefs about gender, power, the reliability of memory, partisan loyalty) on which the audience's priors most diverged. There was no concessionary preamble, no common-ground expansion, no alignment of higher-order assumptions. Evidence was routed through maximally misaligned pathways, and polarisation was the predictable result.

In the language of the formal model: the mediating node M might be "Ford's testimony is reliable," and the prior gap $|h - l|$ between Democrats and Republicans on M was enormous before the hearing began. The hearing provided evidence E (the testimony itself), but because E was routed through M, it widened the posterior gap on P ("Kavanaugh should be confirmed") rather than narrowing it. The CMV money/happiness example shows the alternative: the successful respondent *first narrowed the gap on* M ("you're right that money matters") and only then introduced

evidence that could be processed through the now-shared channel. Nothing in the hearing's structure permitted an analogous move.

What the hearing *lacked* maps onto what the CMV data show works. The most effective CMV replies begin by acknowledging the OP's reasoning, identifying points of agreement, or validating concerns. The formal model predicts that the hearing's structure should polarise, and the polling data confirm it. The lesson is not that the hearing participants were irrational, but that the communicative structure precluded the alignment on which rational convergence depends.

5.5 Limitations

CMV represents a self-selected population engaged under good-faith norms; whether the alignment hypothesis's predictions hold in more adversarial settings is an open question. The strategy labels are LLM-generated with limited human validation (§3.2). The delta badge captures only explicit belief change. I have not yet controlled for confounds such as reply length, position in thread, or topic. What the data currently support is a robust predictive pattern consistent with the alignment hypothesis and not predicted by a simple evidentialist model.

6 Conclusion

I have argued that successful persuasion under conditions of entangled higher-order priors requires alignment on the mediating assumptions through which evidence is weighted. A formal result shows that narrowing divergence on mediating nodes reduces posterior divergence on the target. Large-scale computational evidence from naturalistic debate exhibits the pattern the alignment hypothesis predicts. These results are independently interesting as data about persuasion dynamics, and they resonate with the formal model in ways that reward philosophical attention.

For the epistemology of disagreement, the paper suggests that the conciliationism debate may benefit from structural refinement: what matters is not simply whether to concede, but where in the belief network to concede. For the dialectical/rhetorical divide, it suggests a partial reconciliation: certain rhetorical strategies may serve an epistemic function by creating the conditions under which dialectical reasoning becomes effective. Whether the alignment mechanism also illuminates persistent scientific disagreements (where researchers share data but process it through different methodological priors) is a natural question for future work.

More broadly, the analysis suggests that the gap between what *ought* to persuade and what *does* persuade may be narrower than it appears. If evidence

can only produce convergence when the inferential ground has been prepared, then preparing that ground is not a concession to human weakness but a requirement of rational communication under realistic conditions.

References

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, 4(6):1236–1239.

Blackwell, D. and Dubins, L. (1962). Merging of opinions with increasing information. *The Annals of Mathematical Statistics*, 33(3):882–886.

Christensen, D. (2007). Epistemology of disagreement: The good news. *The Philosophical Review*, 116(2):187–217.

Clark, H. H. and Brennan, S. E. (1991). Grounding in communication. In Resnick, L. B., Levine, J. M., and Teasley, S. D., editors, *Perspectives on Socially Shared Cognition*, pages 127–149. American Psychological Association.

Elga, A. (2007). Reflection and disagreement. *Noûs*, 41(3):478–502.

Freeborn, D. P. W. (2023). *Polarization and Factionalization for Agents with Multiple, Related Beliefs*. PhD thesis, University of California, Irvine.

Freeborn, D. P. W. (2024). Rational factionalization for agents with probabilistically related beliefs. *Synthese*, 203(2):1–27.

Jern, A., Chang, K.-m. K., and Kemp, C. (2014). Belief polarization is not always irrational. *Psychological Review*, 121(2):206–224.

Kelly, T. (2008). Disagreement, dogmatism, and belief polarization. *The Journal of Philosophy*, 105(10):611–633.

Marist Poll (2018a). NPR/PBS NewsHour/Marist Poll National Tables, October 2018. Marist College Institute for Public Opinion.

Marist Poll (2018b). NPR/PBS NewsHour/Marist Poll National Tables, September 2018. Marist College Institute for Public Opinion.

Mercier, H. and Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2):57–74.

Sicilia, A., Kim, H., Chandu, K. R., Alikhani, M., and Hessel, J. (2024). Deal, or no deal (or who knows)? Forecasting uncertainty in conversations using large language models. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 11700–11726. Association for Computational Linguistics.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., and Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, 25(4):359–393.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, 25(5–6):701–721.

Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., and Lee, L. (2016). Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web*, pages 613–624.

van Eemeren, F. H. and Grootendorst, R. (2004). *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*. Cambridge University Press.

Walton, D. N. (1995). *A Pragmatic Theory of Fallacy*. University of Alabama Press.